# The Cell and its Blueprint

The cell is the functional unit of life. This concept is established in what we call "Cell Theory," which was originally proposed in **1839** by Matthias Jakob Schleiden and Theodor Schwann. While identifying cells as as fundamental to life was a major step in biology, cell theory lacked a description of what makes a cell. In other words, what is the blueprint of a cell? Today we know the answer to this question: DNA. However, for over a century after the description of Cell Theory scientists argued about the molecular blueprint that stored cellular information. In this section we will walk through the several major breakthroughs in the history of our understanding of DNA.

# The Discovery of DNA

DNA was originally discovered by Johann Friedrich Miescher, a Swiss chemist who collected puss-covered bandages from medical clinic patients in **1869**. Friedrich added salt and acid to the bandages to see what would happen, and in doing so he isolated a substance he called "nuclein" (believing it came from the cell nucleus). Today we know nuclein as DNA. The gravity of this finding was not understood, especially considering that most scientists at the time believed proteins to be the molecules of heredity.

| Substitution Point Opportunity:<br>Isolate "nuclein" just like Johann Friedrich Miescher did it.<br>See the instructions on Canvas under "Assignments → Substitution Point Opportunities → DNA Isolation" |
|---|

Johann Friedrich Miescher was mentored by Albrecht Kossel, a German biochemist who was interested in the nuclein discovered by his student. Upon investigating the components of nuclein, in **1881** Kossel learned that it was together in the nucleus with proteins. Kossel described the five nucleotide bases (Adeneine, Cytosine, Guanine, Thymine, and Uracil) in **1888**. He was awarded the Nobel Prize in Medicine/Physiology in 1910. Today we call the combination of DNA and proteins "**chromosomes**", and we call the proteins (which DNA wraps around) "**histones**.

# The Molecule of Inheritance

**Inheritance** is the process of passing biological information from one generation to the next. Something that is inherited through generations is called "**heritable**." **Heredity**, therefore, is the simply the concept of inheritance. **Genetics**, at its core, is the scientific study of heredity, however during this semester you'll see that genetics expands into the study of genes, genotypes, and phenotypes.

The first people to propose chromosomes as the molecules of heredity were Walter Sutton (a scientist at the University of Kansas who played basketball under coach James Naismith) and Theodor Boveri (a German scientist who saw science as a combination of nature and art) around 1902. Their theory became known as the **Sutton-Boveri Chromosomal Theory of Inheritance** (however, DNA was not yet invoked as the molecule of heredity– remember, chromosomes were made up of both DNA and proteins). Sutton and Boveri didn't know each other, but they had two things in common: (1) they both were undecisive with their undergraduate majors

(Boveri switched from history/philosophy to anatomy/biology; Sutton switched from engineering to biology to medicine) and (2) they independently suggested that chromosomes possessed the information required for the heritability of traits. Boveri described "reduction division" (a term you will learn about in the Meiosis section), and he noted equal contribution of sperm and egg cells. Sutton wrote "The association of paternal and maternal chromosomes in pairs, and their subsequent separation during the reducing division... may constitute the physical basis of the Mendelian law of heredity." This brings up two new words that will be the focus of the next section: "Mendelian" and "heredity".

# The Discovery of Heredity

An understanding of heredity as a principle started waaaay before Sutton and Boveri. In fact, we can't even identify when heredity was "discovered". It is logical to assume that independent groups recognized that traits were passed from one generation to the next. For example, the cultivation of maize from wild teosinte to what we call corn occurred **8,000 B.C.E.** and required an understanding of heredity (cultivators knew that by using seeds from plants with "yummier" corn they would end up with more "yummy" plants). Hippocrates, the ancient Greek known today as the "Father of Medicine", proposed **400 B.C.E.** that "All parts of the body give seed that are transmitted during sexual intercourse" (an idea known as **pangenesis**). In **1859** Charles Darwin based his entire mechanism of evolution by natural selection on the idea that traits in offspring are inherited from parents. However, Darwin also invoked pangenesis to explain the heredity, and suggested "**blending inheritance**" (wherein traits of two different parents blended in subsequent generations) as the mechanism of inheritance.

The exact nature of heredity wasn't well described until **1865** when a Czech monk named Gregor Mendel presented an article titled "Experiments in Plant Hybridization" to a small science group called *The Nature Research Society of Brno*. Mendel meticulously examined generations of garden peas to understand how a trait was passed from parent to offspring, and he made a major breakthrough by discovering that the underlying elements of heredity were discrete structures he called "elemente" (structures that today we call "**genes**"). Today we call unique versions of a gene **alleles** (Mendel called them "merkmale"). The idea of discrete, variable elements rejected the idea of "blending inheritance" proposed by Darwin, because the discrete nature of genes means they could never be "blended"– they were either present or absent. Mendel found that when he crossed plants with unique versions of a trait (e.g., a plant with purple flowers with a plant with white flowers), that 25% of all the offspring exhibited one version (e.g., white flowers) and the other 75%. had the other version (purple flowers). This 3:1 ratio was consistent across 7 different traits he examined, and led him to develop the "**law of segregation of alleles**". In a similar experiment where he looked at two traits together (e.g. flower color and plant height), Mendel noticed that the parental versions of traits weren't inherited together (i.e., offspring from a cross between tall purple plants and short white plants could be (a) tall and purple, (b) short and white, (c) tall and white, or (b) short and purple). This led him to develop the "**law of independent assortment**". These two laws are foundational to the field of genetics, and because of this Mendel is commonly referred to as the "Father of genetics". Despite the magnitude of this scientific breakthrough, Mendel lived and died without his work being noticed by other scientists. Knowledge of the laws of heredity would have filled the major gap in Darwin's theory of natural selection, yet after Darwin's death a letter from Mendel, who didn't know Darwin on any personal level, was found unopened in Darwin's library. We will discuss the laws of genetics discovered by Mendel, along with the details of his experiment, in a future section.

Mendel's work describing the discrete nature of heredity and the two laws of genetics went unnoticed for nearly 50 years, until it was independently re-discovered in **1900** by three scientists (Hugo De Vries [Netherlands], Karl Correns [Germany], and Erich von Tschermark [Austria]). While each of these recognized Mendel as the originator of the principles of discrete heredity, none of them promoted Mendel as much as the English zoologist **William Bateson**. Bateson was nicknamed "Mendel's Bulldog" because he was so adamant about promoting "Mendelism" (the idea of heredity through discrete elements) and establishing Gregor Mendel as the figurehead of a new field Bateson called "**genetics**" in **1905**. Mendelism surprisingly received significant pushback from many (if not most) prominent biologists interested in heredity, but the persistence of scientists such as Bateson to convey the evidence of Mendelian genetics finally convinced the general field of scientists that heredity occurs via discrete units (genes). In addition to coining the term "genetics", Bateson also coined the terms "zygote", "homozygote", "heterozygote", and "allelomorph" (today we just say "allele"). Bateson (with his assistant, Reginald Crundall Punnett) also created the useful technique to visualize genotype probability that we call the "Punnet square".

Interestingly, the term "gene" was coined (in **1909**) after the term "genetics" by the Danish scientist **Wilhem Johannsen**. Johannsen also coined the terms "genotype" and "phenotype". Additionally, Johannsen discovered that organisms with the same **genotype** (genetic composition) could still have a different **phenotype** (version of an observable trait). Today we call this concept **phenotypic plasticity**, which is when the environment can produce unique phenotypes in organisms that have the same genotype.

While the idea of genes as the elements of heredity had become established, they were still obscure entities with no ties to any physical components in the cell. As mentioned before this section, Sutton and Boveri were the first to propose chromosomes as the substance of Mendelian heredity. **Thomas Hunt Morgan**, a biologist at Columbia University in New York, piggy-backed of of their idea saying that genes were located on chromosomes just like beads on a string. This was a big transition for Morgan, who was originally a staunch opposer of Mendelian genetics as the explanation of heritability of traits. He would go on in **1911** to prove that genes were located on chromosomes using *Drosophila* (vinegar flies) as his model system. Morgan also proposed "crossing over" as a mechanism to explain independent assortment of genes located on the same chromosome, a concept that would not be proven for over 20 years. Morgan was awarded the Nobel Prize in Physiology or Medicine in 1933. Recombination was later proven by Cornell geneticist **Barbara McClintock** in 1930. McClintock would later make breakthrough discoveries involving telomeres, centromeres, genetic transposition, and genetic mapping. She was awarded the Nobel Prize in Physiology or Medicine in 1983.

# Eugenics

During the scientific buzz surrounding the genetic enlightment in the early 1900s was a looming darkness. Eugenics, a practice that promoted reproduction among people with "favorable" traits and inhibited (or outright prevented) reproduction among people with "unfavorable" traits (such as feeblemindedness, epilepsy, criminality, insanity, alcoholism, pauperism, among others) had become a prominent topic in the United States. This deplorable practice was supported by the majority of geneticists in the United States, many of whom misconstrued and twisted principles of genetics to try to fit them within the eugenics agenda. Morgan and Bateson were among those who opposed the practice based on the fact that it lacked any scientific foundation and

it targeted oppressed groups. The practice of forced sterilization implemented by the American genetics movement eventually inspired the genocide of diverse groups at the hands of Nazi Germany.

## DNA as the Molecule

In **1952 Alfred Hershey** and **Martha Chase**, geneticists in Cold Spring Harbor New York, proved that DNA is the molecule of inheritance. They radio-labeled atoms of proteins (sulfur) and atoms of DNA (phosphate) within a virus that infected bacterial cells to investigate whether protein or DNA was used to replicate the virus. At the time there was still active debate concerning whether proteins or DNA served as the molecules of inheritance. Hershey and Chase's experiment proved that the DNA was transported into the cell, and it was then used to create novel virus. To better understand how this experiment determined that DNA is the molecule of heredity, check out this video: `https://www.youtube.com/watch?v=ZtSfFqqhEIY`. Hershey was awarded the Nobel Prize in 1969, Chase was not.

During this same time, Rosalind Franklin was seeking to understand the structure of DNA at King's College in England. Franklin used X-ray crystallography to decipher structures of complex molecules and compounds. She published her famous image called "photograph 51" in an article in the journal *Nature* titled "Molecular Configuration in Sodium Thymonucleate" in **1953**. It is a common misconception that she never published her work. In fact, this article was published in the same year in the exact issue of the same journal where James Watson and Francis Crick published their article titled "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid". In this article, Watson and Crick describe the antiparallel double helix - the structure of the DNA molecule. It is important to recognize that a polymer is a collection of linked/bounded monomers (i.e., think of a polymer as a "brick wall" and monomers as the "bricks"). The monomers of DNA are nucleotides. Each nucleotide contains (1) a phosphate group, (2) a five-carbon sugar, and (3) a nitrogenous base. There are four nitrogenous bases in DNA: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). The DNA double helix is made up of two strands of nucleotides- one strand is bound at its backbone by covalent bonds, and the two strands are bound to each other with "weak" hydrogen bonds. To better understand the structure of DNA, check out this youtube video: `https://www.youtube.com/watch?v=o_-6JXLYS-k`. Controversy surrounds these publications, as Watson and Crick discovered the structure of DNA with the help of photograph 51, which was sent to them from a colleague (Maurice Wilkins) that took it from Franklin's desk without her permission. Crick, Watson, and Wilkins were awarded the Nobel Prize in Physiology or Medicine in 1962, just four years after Franklin's death.

## The Central Dogma

Before publishing the description of the double helix, Watson drew a diagram of "DNA $\rightarrow$ RNA $\rightarrow$ Protein". Crick later called this idea the "Central Dogma" of biology. He used the term "dogma" in its religious context, because he felt there was no direct evidence for the idea but it was powerful. Essentially, he was mocking religion and claiming the importance of this idea in a single word- and it stuck.

The central dogma can be summarized in Figure 1. The first component of the central

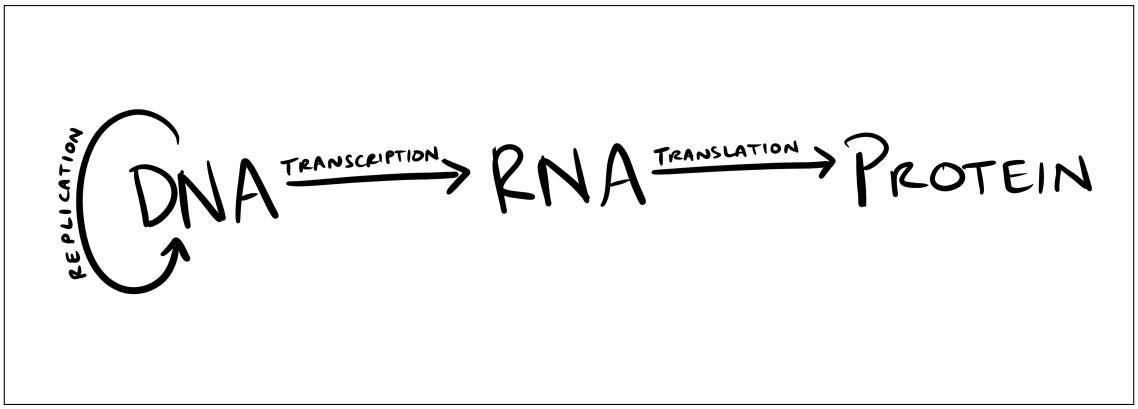


Figure 1: Central Dogma of Biology

dogma is that DNA **replicates** (copies itself). The mechanism of replication was discovered by Harvard biologists **Matthew Meselson** and **Franklin Stahl** in what has been called "the most beautiful experiment in biology". Meselson and Stahl took DNA that contained the $N_{15}$ isotope and surrounded it with $N_{14}$ nucleotides. To discover how DNA replicated, Meselson and Stahl measured the DNA after cycles of replication (using a centrifuge). Before any replication occurred, all of the DNA was located towards the bottom of the column (which made sense- all of the DNA contained the $N_{15}$ isotope). After one round of replication, there was no longer a band at the previous location; all of the DNA was slightly lighter. This is not what would be expected if the DNA replicated via "conservative" replication (where an exact copy of the original strand was made without breaking apart the original strand). If conservative replication was the mechanism, then two bands should have been visible (one at the $N_{15}$ position and another at the $N_{14}$ position). However, Meselson and Stahl observed a single band between where they saw the original $N_{15}$ band and where they would have seen a $N_{14}$ band. This indicated that the double stranded DNA was a hybrid of $N_{15}$ and $N_{14}$ DNA, which is what would be expected if the replication process was semiconservative. In semiconservative replication, the original strand is "unzipped" at the hydrogen bonds that join the two strands. Each subsequent strand is then used as a template to synthesize the creation of a new strand. The final result is two double-stranded DNA molecules, each of which contains a strand with $N_{15}$ and a strand with $N_{14}$. The next round of replication resulted in a new band at the $N_{14}$ location, while the band between the $N_{15}$ and $N_{14}$ was still there. This is a result of the hybrid (half $N_{15}$, half $N_{14}$) strands undergoing their own replication, which means the resulting DNA will include molecules with only $N_{14}$ and other molecules that are hybrid $N_{15}/N_{14}$. To better understand this experiment, check out this youtube video: `https://www.youtube.com/watch?v=4gdWOWjioBE`.

The next step of the central dogma is transcription. While Watson and Crick suggested the idea of DNA being converted to protein using an RNA intermediate, Meselson published on the concept in a paper titled "An unstable intermediate carrying information from genes to ribosomes for protein synthesis". DNA in the nucleus is protected from the mutagens in the cytosol, but this presents a problem if DNA is the blueprint: how are the builders supposed to read the instructions if they are locked away? RNA is the solution- these molecules are temporary copies of the DNA that can be sent into the cytosol to provide information to the builders. The process of creating this temporary copy is called **transcription**- an exact copy of the DNA sequence is created. Both DNA replication and transcription require an enzyme to synthesize a new nucleic acid (be it DNA or RNA). This enzyme is known as a **polymerase** (think of it as something that creates a polymer). The discovery of DNA polymerase in **1956** by Spanish molecular biologist **Severo Ochoa** and U.S. molecular biologist **Arthur Kornberg** resulted in their receiving the Nobel Prize in Physiology and Medicine in 1959.

However, something needs to be able to interpret the code provided by RNA and convert it into proteins. If we compare DNA to a written language, you can think of the bases (A, C, G, and T) as the "letters" of the language. However, as you are aware letters alone don't make a language- words (delineations between letters) allow for us to interpret letters. In **1961** U.S. scientists **Robert Holley**, **Har Bogind Khorana**, and **Marshall Nirenberg** discovered that the "words" of the genetic code were three-base long sequences they called **codons**. In **1965** They discovered that each RNA codon (combination of three nucleotide bases A, C, G, and U [Uracil replaces Thymyine in RNA) encodes an amino acid. The site of translation is the ribosome, where tRNA molecules carrying amino acids are matched to codons (the molecular details of this process will be discussed later in the semester). These researchers were awarded the Nobel Prize for Physiology or Medicine in 1968. To better understand the processes of transcription and translation, check out these videos: `https://www.youtube.com/watch?v=oefAI2x2CQM`, `https://www.youtube.com/watch?v=LsEYgwuP6ko`.

# The Birth of Biotech

Soon after the nature of the central dogma was established, researchers began to explore what we could do with this newfound knowledge. In **1972** U.S. geneticists **Herbert Boyer** and **Stanley Cohen** created the first recombinant DNA. They inserted a frog gene into a bacterial plasmid. By inserting this plasmid into a bacterial cell, they could utilize the bacterial cell machinery to replicate the plasmid (a process called "recombinant cloning") and transcribe + translate the protein they had inserted.

While recombinant cloning was a major breakthrough in genetics, obtaining the genetic sequence (i.e., the arrangement of nucleotide bases) within a gene was still a challenge. In **1977** English molecular biologists **Walter Gilbert** and Frederick Sanger used di-deoxynucleotides the concept of gel electrophoresis to create a way to visualize the exact genetic sequence of a gene. This technology is known as "**Sanger sequencing**" today, and Gilbert and Sanger were awarded the Nobel Prize in Chemistry in 1980.

However, obtaining enough copies of a gene through recombinant cloning was a painstaking process. In **1985** the Californian biologist Kary Mullis working at the biotech company Cetus created the **polymerase chain reaction (PCR)**. This idea utilized heating/cooling cycles to **denature** (separate the DNA at the central hydrogen bonds), single-stranded DNA molecules called "primers", and a heat-tolerant polymerase to copy targeted genetic regions. Mullis won the Nobel Prize in Chemistry in 1993.

# Genomics

The techniques of Sanger sequencing and PCR were used to sequence the first human genome, an endeavor that took over a decade and 3 billion U.S. dollars. Two groups (Celera [a private company] and the Human Genome Consortium [a U.S. federally funded group]) completed the sequencing of the human genome at the same time, publishing their findings in the journals *Science* and *Nature* in the same month of **2001**. Following the conclusion of the human genome project, the field of genomics exploded. Bioengineering breakthroughs resulted in the development of machines that could perform sequencing at a massive scale through a technique

known as "**Next-Generation Sequencing (NGS)**". The major companies driving this growth include illumina, Pacific Biosciences (PacBio), Roche 454, and Oxford Nanopore. Because of these breakthroughs, today a human genome can be sequenced for less than $1,000 U.S.

Increased accessibility to genomic sequencing expanded the scope of genomics into the fields of physiology, ecology, evolution, epidemiology, conservation biology, forensic biology, and others. The number of gene sequences archived on the U.S. database Genbank has increased almost exponentially due to the feasibility and utility of genomic sequencing. The Nobel Prize in Physiology or Medicine for 2022 was awarded to the Swedish geneticist and paleobiologist Svante Paabo, who in **2010** sequenced the Neanderthal genome and discovered historical gene flow between *Homo sapiens* (our species) and *H. neanderthalensis* (neanderthals). Molecular biologists **Jennifer Doudna** and **Emmanuelle Charpentier** discovered CRISPR from the genomes of bacterial cells in **2012**, which information they used to create a hyper-specific genomic editing tool (CRISPR-Cas9) and earn the Nobel Prize in Chemistry in 2020. And in **2020** molecular biologists Drew Weissman and Katalin Kariko harnessed the cells translational machinery to create a mRNA vaccine based on the genome of the SARS-CoV-2 genome, for which they were awarded a nobel prize in 2023. The field of genetics is progressing at a remarkable pace, and the effect of these advances on society are evident. From basic research to medicine to agriculture, we can see the impact of modern genetics almost everywhere we look. What will be the next big breakthrough?